

Detecting Newcomb-Benford Digital Frequency Anomalies in the Audit Context: Suggested χ^2 Test Possibilities

Edward J. Lusk^{1,2} & Michael Halperin³

¹The State University of New York (SUNY) at Plattsburgh, Plattsburgh, NY, USA

²Emeritus, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

³Director Lippincott Library of the Wharton School, University of Pennsylvania, Philadelphia, PA, 19014, USA

Correspondence: E. Lusk, SBE SUNY Plattsburgh 101 Broad St. Plattsburgh, NY, USA 12901, Tel: 1-518-564-4190
E-mails: lusk@plattsburgh.edu or lusk@wharton.upenn.edu

Received: April 22, 2014

Accepted: May 12, 2014

Online Published: May 20, 2014

doi:10.5430/afr.v3n2p191

URL: <http://dx.doi.org/10.5430/afr.v3n2p191>

Abstract

Digital Frequency Testing [DFT] has achieved justifiable currency as a valuable part of the panoply of the auditor. There are a number of inference models, such as the parametric test for proportional differences to entropic screening, that can be used to create information regarding the use of extended procedures to investigate the difference between the *Observed* digital frequency compared to the *Expectation* benchmark. One inferential model which seems ideal for DFT in the audit context is the Chi2 model as it is formed as the ratio of the square of the difference between the *Observed* and the *Expectation* benchmarked by the *Expectation*. This Chi2 ratio fits perfectly the usual audit investigation sensitivity imperative developed by the analytic procedures phase of the audit. However, there are *Expectation Specification* and *Sample Size* issues that need to be considered in using the Chi2 test in DFT. We present and illustrate in detail three testing models focusing on the Pearson and the Direct Benchmarking expectation differences with suggestions as to sample size control that could be used by the auditor in the certification audit. We offer important *audit action information* regarding the inference information that is related to the overall Chi2 test as well as to the Chi2 cell contribution signals. Finally, we have a Decision Support System [DSS] that aides in the creation of this Chi2 audit-screening information. The DSS, programmed in Excel-VBA, is available from the corresponding author as a free download without restriction to its use.

Keywords: Benford and Log10 Expectation Projections, Chi2 Two Sample Model

1. Introduction

Ever since the passage in 2002 of the sorely and surely needed Sarbanes-Oxley:[SOX] US Federal legislation [(Pub. L. 107–204, 116 Stat. 745)] that created the Public Company Accounting Oversight Board [PCAOB], there has been a commitment to monitoring information reported by firms listed on trading exchanges. See *The Report on Financial Oversight of Enron* (2002). The logistics of this commitment are effected through AS 5 (PCAOB, 2007), which are the rules by which the PCAOB judges the effectiveness of the certification audit. The chain of responsibility PCAOB→AS 5→Certified Public Accountant performing the audit has the intention of providing reasonable assurance that the information published in the annual report/10-K by trading organizations is free from material error, and is therefore a reasonable reflection of the results of their operations reflected through the GAAP-USA lens. The crafting of AS 5, which replaced the initial set of audit regulations promulgated as part of SOX [AS 2], benefited from extensive input from the SEC, the AICPA, practicing Public Accountants and the research published in peer reviewed journals. The execution of the audit envisaged by the PCAOB is dynamic and changes over time through the issuance of pronouncements, opinions, research, and interpretations that help clarify the meaning of “best practices” in the audit context. This dynamic evaluation of the best practices profile interestingly is not expressed in any official document. *It is subjective but objectively applied by the PCAOB and other regulators*; what was considered best practices in the early 1980s may not now be judged as adequate audit investigation to deliver on the assurance consistent with AS 5 with which the Certified Public Accountants are charged. The dynamic evolution of the best practices concept is the point of departure of this paper.

In the vast milieu of:

- AIS-software, audit programs and protocols detailing tens of thousands of possible tests directed at the various conceptualizations of inferentially relevant audit information,
- the numerous versions of the *Management Assertions* that are generally accepted by the PCAOB, the SEC and CPAs as the foci or the GPS of the audit,
- powerful Audit GUI IT interfaces, such as the widely used IDEA™ audit software [<http://www.caseware.com>] and the related statistical sampling protocols, and
- the innovative investigation perspectives such as Content Analysis, for a recent example see: Lee, Lusk & Halperin, (2014) and Benford Digital Frequency-profiling,

we wish to offer a discussion of the Benford profiling using the χ^2 category test as the inferential model regarding the use of extended procedures in certification audit.

Specifically, we will: (1) detail and motivate the use of the Benford Profile in the audit context, (2) examine the theoretic and the practical aspects of using the χ^2 as the inferential model for detecting Benford digital frequency anomalies, and finally, (3) suggest, in illustrative detail, possible uses of the χ^2 distribution in the important decision as to whether extended procedures *are* or alternatively *are not* warranted in the context of the certified audit.

2. The Benford Profile

One important inferential signal in deciding on the use of extended procedures testing for an account under audit regarding errors caused by incompetence, inattentiveness, or “localized” defalcations/fraud can be generated by examining the distribution of the frequency of digits in the recorded account. For example, assume the auditor received from the client, a large retail department store such as *Walmart*, a download of the year-end retail prices of the ending inventory and partitioned this inventory “population” into two price categories: goods priced under \$100 and those priced at or over \$100 so as to execute a stratified random sampling protocol. Consistent with GAAS: General Standard 3: *Due Professional Care* assume that the auditor scanned recorded prices of this list as a preliminary “reasonability check”. If the auditor observed that the percentage of the items priced \$xx.x9 were to be 9.8%, this would likely be a noteworthy deviation from the usual expectation that: *Almost all retail items priced under \$100 usually end in \$xx.x9*. This variance [expecting most of the items to be priced \$xx.x9 but finding only 9.8% of the inventory items were so priced] would likely require the auditor to inquire as to the nature or reason(s) of this divergence of *Actual* from *Expectation*. Depending on the explanations offered by management extended procedures *may* or *may not* be warranted. For example, assume management indicated that based upon consumer focus-group interviews conducted by the marketing department it is possible to create positive product differentiation by pricing all items so that there were random digits *in the unit dollar and cents places* [\$x.xx]. After reviewing this marketing department plan, what the auditor observed, 9.8% of the accounts ending in \$xx.x9, would likely be judged as consistent with the pricing policy and so extended procedures would not be warranted in testing this partition of the inventory account. This observation, subsequent inquiry of *Walmart’s* management, and the final audit disposition would be written up by the auditor and included as part of the current audit record which then would become part of the permanent file of the engagement. Therefore the digital frequency profile [DFP] of audit accounts and transaction sets is one aspect typical of audit evidence of interest in the best practice execution of the audit relative to rationalizing the decision regarding electing to use extended procedures in conducting the audit.

This being the case, auditors have discovered how to use the DFP observation of Newcomb (1881, p. 39)

“That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones.”

Later Benford (1938) reported the same observation. We refer to this as the [N-B Profile]. Using this empirical evidence Newcomb and Benford both arrived at a simple formula to characterize the likely distribution of the nine first digits. To wit:

$$\text{Frequency}[d_i] = \text{LOG}_{10} (1 + 1/d_i) \text{ for } i= 1, 2, \dots, 9 \quad (1)$$

Therefore, this simple formula, if it is, indeed, the underlying generating process for digital frequencies in the economic milieu, can be used by the auditor to benchmark particular *Observed* digital frequency profiles from accounts under audit examination. For example, assume that the auditor has this same population of Inventory items. For the partition of items priced over \$100, the auditor now can use EQ1 to generate the *theoretical* digital profile of these individual items; using this *Expectation* as the benchmark, the auditor can examine the resulting *Actual* digital profile in comparison. In this regard, the Digital Frequency Profile [DFP] expected according to the N-B theoretic

expectation for the percentage of items in this partitioned Inventory-set having first digits of “1” or “2” would be: N-B Benchmark[Digits: 1,2]: $\text{LOG}_{10}(1 + 1/1) + \text{LOG}_{10}(1 + 1/2) = \text{LOG}_{10}(2) + \text{LOG}_{10}(1.5) = 30.1\% + 17.6\% = 47.7\%$

If the auditor observes that the percentage of Inventory items starting with "1" or "2" were to be 45.2%, a difference between *Observed* and *Expected* of: -2.5% [$45.2\% - 47.7\%$] would likely be accepted as being in the usual random variation experiential range. In this case, the auditor would likely not feel that extended procedures were to be warranted.

Most auditors would agree; however, the PCAOB best practices imperative requires more than just a “feeling” for the justification of deciding not to use extended procedures; the auditor needs to have an objective inferential justification. The fact the auditor “felt” that the difference between the *Observed* and *Expectation* was relatively small and so made the decision not to use extended procedures is a personal judgment that may at some point be called into question. It would have been much better if the auditor had an objective measure suggesting that: *A difference of -2.5% between the Observed and Expected digital frequency profiles is sufficiently small so as not to provide sufficient evidence to reject the belief that such a difference could have resulted from random sampling chance for a population where the first two first digits constitute 47.7% of the first digits.* This is, as one recognizes, the *Null* of no difference which in this case cannot be rejected with a reasonable degree of certainty given the small test against difference of -2.5% . Without such a False Positive Error [FPE] inferential testing context there is a certain jeopardy for the auditor to make the call to not use extended procedures based upon a feeling unsubstantiated by objective inferentially valid evidence. Therefore, in the audit testing context there needs to be an objective inferential measure relative to the benchmark to justify the decision *to use* or *not to use* extended procedures. We have been using the N-B Log_{10} formula as the benchmark as it is the case that most researchers (Note 1) have identified this formula as a reasonable test-against benchmark. However, recent research has offered another possibility. Consider now this alternative benchmark.

2.1 An Alternative Digital Frequency Benchmarking Profile

This alternative benchmark is due, in fact, to Benford. To give operational validity to the Log_{10} generating function, Benford (1938, Table 1, p. 553) collected 20 samples from an impressive spectrum of generating processes, such as: *River Drainage Areas*, *Economic Costs*, and *Atomic Weights* to mention a few. The number of observations, in total, for these 20 datasets is 20 229. The range of the sample sizes for the 20 accruals is [91 to 5000] with a mean of 1012. Therefore, these frequencies as “a realization-profile” could also be used as a benchmark for the *Observed Digital Frequency* profile. However, due to recent research of Lusk and Halperin (2014b), it was reported that the mean frequency profile reported by Benford (1938, Table 1, p. 553) may be refined. Lusk and Halperin, offer:

“Benford reports a curious measure: In Table 1, p. 553 we find the Average [Arithmetic Mean] for the 20 point measures for the first digit set. To produce this measure, he took the simple average of the point frequency measures as reported in Table 1. Actually, this has no useful statistical meaning. The correct frequency average is the number of digits in a particular first digital Bin as a ratio to the sample total of 20,229. For example, in Table 1, p.553 Benford has the frequency proportion average for “1” as: 30.6%. However, the correct frequency percentage is the unit-vector product of the 20 frequencies with the respective sample sizes for the first digit. This develops for the 20,229 observations that 5,849.295 were “1s” and the proper ratio is $5,849.295/20,229 = 28.9\%$. This is essentially the weighted average Mean, which in this case is the actual frequency. For Windowing or Screening purposes, this orientation anomaly created by using the simple average could be important in profiling actual datasets. Therefore, with apologies to M. Benford we corrected his average calculations. In Table A are the corrected Mean frequencies and a suggested screening window for the comparative evaluation of collected audit datasets using this corrected information.”

This suggests that for the Newcomb-Benford profile that the auditor could use the Log_{10} generating function or could use the Lusk & Halperin (2014b) corrected means for the 20 samples that Benford reports termed the Benford Practical Profile [BPP]. These are presented in Table 1 following:

Table 1. The Benford Practical Profile [BPP] and the N-B Log₁₀ Profiles

First Digit Array	Corrected Means of The B-N Log ₁₀ Digital Difference [Col2 less
	Benford Datasets: BPP Profile Col3] Values
Digit 1	0.289 19 0.301 03 -0.011 84
Digit 2	0.194 62 0.176 09 0.018 53
Digit 3	0.126 65 0.124 94 0.001 71
Digit 4	0.090 61 0.096 91 -0.006 30
Digit 5	0.075 44 0.079 18 -0.003 74
Digit 6	0.064 31 0.066 95 -0.002 64
Digit 7	0.054 08 0.057 99 -0.003 91
Digit 8	0.054 87 0.051 15 0.003 72
Digit 9	0.050 52 0.045 76 0.004 76

These two benchmarks are effectively, and certainly not surprisingly, substantially similar; for example, the sum of the differences in Col 4 is 0.000 29 and the distribution of the signs is as equal as is possible. However similar these two digital profiles appear to be *in the aggregate*, individually, as orientations in Cartesian coordinate space, they are non-trivially different and so may be considered to be different conceptual benchmarking variables. Specifically, the average of the absolute value of the differences in Col 4 as a ratio to the average of the two benchmarks is 6.1%. This leads us to the next contextual issue: The nature of the assumption underlying the development of the χ^2 as an extended procedures screening protocol.

3. The χ^2 test: An Ideal Conceptual Model

As indicated above, we suggest that the χ^2 Test for Digital Frequency Benchmarking fits very well in the audit context as the χ^2 test is *Observation* as it varies from *Expectation*; however, the χ^2 test has not achieved currency in the audit context for two practical reasons: (i) The Expectation or Benchmarking Election and (ii) Sample Size issues. We will consider these next and suggest methods of data testing that will enable the auditor to use the χ^2 test in the audit context.

3.1 The Expectation Selection Issue in the Audit Context

There are two ways that the *Expectation*, whether the BPP or the classic N-B Log₁₀ formula, used in making the χ^2 computation may be integrated into the computation: The **Two-Sample Category** computation due to Pearson (1900) and the **Direct Benchmarking** due to the form of the χ^2 computation. The nature of these computations usually creates significant inferential differences, and so can change the action plan of the auditor. To illustrate the inferential effects of these two computational forms, essentially for either the BPP or the N-B Log₁₀ formula, we will now consider the formulae for The Pearson and the Direct Benchmarking.

3.1.1 Two-Sample Category Pearson Formulation

The classic Pearson (1900) formula in the N-B context for the χ^2 is:

$$\chi^2 = \sum_{j=1}^2 \sum_{i=1}^9 [(O_{ij} - E_{ij})^2 / E_{ij}] df = 8, \quad (2)$$

Where: O_{ij} represents the observed frequency/number for i^{th} digit for the j^{th} sample, E_{ij} represents the expected frequency/number for the i^{th} digit for the j^{th} sample. For example, if there are two sample events or category profiles there are 9×2 or 18 values in the summation for EQ2.

The logic underlying EQ2 is that there are two samples: One from the audit—a digital frequency profile of the account under audit, for example: $j = 1$ and a sample from the category benchmark as the projection of the DFP taken from the corrected frequencies of the Benford samples or the Log₁₀ formulation where then $j = 2$. Using the Pearson formulation in the DFT context, the Expectations, E_{ij} , would be computed as the simple average of the relative marginal values. This is the formulation due to Pearson (1900) for comparing two sample realizations. As the total number of observations in each of the two DFP samples is identical, E_{ij} for the audit benchmarking is simply computed as:

$$E_{ij} = (\sum_{j=1}^2 O_{ij}) \times 50\%, \text{ over } i: 1, 2, \dots, 9. \quad (3)$$

For example, assume that we have 850 observations as the population of Accounts Payable GJL balance sheet accounts from the GIC, Sector 25: WRDS™. If we wish to use the Benford Practical Profile [BPP] as the sample benchmark then we will create these nine digital frequencies, df_i , as follows:

$$df_i = (BPP_i) \times N_i: 1,2, \dots, 9. \quad (4)$$

Where: BPP represents the Benford Practical Profile as taken from Table 1 Col2, and N is the number of observations in the population of Accounts under audit.

Applying EQ4, generates the following values as found in Table 2:

Table 2. Benchmarking the A/P Datasets Using the BPP

First Digit Array	Corrected Means: Benford Practical Profile [BPP]	A/P Benchmark using the BPP:EQ4	Actual Digital Counts for the A/P
Digit 1	0.289 19	245.81	262
Digit 2	0.194 62	165.43	156
Digit 3	0.126 65	107.65	98
Digit 4	0.090 61	77.02	91
Digit 5	0.075 44	64.12	67
Digit 6	0.064 31	54.66	49
Digit 7	0.054 08	45.97	44
Digit 8	0.054 87	46.64	44
Digit 9	0.050 52	42.94	39
Totals	N/A	850.25	850

As one can see the sample sizes are effectively the same for the usual rounding to the nearest integer.

3.1.2 Continuing with the Pearson Computation of the χ^2 for the two sample comparison

For the A/P benchmarked against the Benford Practical Profile—i.e., referencing EQ2 and EQ3—we have the intermediate calculation information as presented in Table 3:

Table 3. Computation of Expectations using the Marginals in the Pearson Formulation

Digits	BPP Projections Counts	A/P Counts	Marginal Values	Expectations (Note 2)
1	245.81	262	507.81	253.91
2	165.43	156	321.43	160.71
3	107.65	98	205.65	102.83
4	77.02	91	168.02	84.01
5	64.12	67	131.12	65.56
6	54.66	49	103.66	51.83
7	45.97	44	89.97	44.98
8	46.64	44	90.64	45.32
9	42.94	39	81.94	40.97
Marginals	850.25	850	1700.25	850.12

In this case, the expectations can be derived in two ways: the Pearson Marginal products, related to the sum of the two data samples or the formula: EQ3. The two are essentially always the same in the projection case treated above. For example, consider the expectation for digit 2 for the BPP [BPPExp(d_2)]. The Pearson Formulation is:

$$BPPExp(d_2) = [Marginal(BPP) \times Marginal(d_2)] / GrandSum \quad (5)$$

$$BPPExp(d_2): 160.7 = \frac{(850.25 \times 321.43)}{1700.25} \quad (6)$$

As we see from EQ 3 above $160.7 = 321.43 \times 50\%$. This is immediate from the above Pearson formulation as the Column Marginal count values [BPP and A/P] are essentially equal [at 850 each] they will be 50% of the Grand Sum. Therefore, for the Two Sample Pearson case, 50% of each of the nine Row marginals will be the expectation for both datasets as presented in EQ3. Further, the 18 individual row χ^2 cell contributions will also be the same which are summed to form the overall χ^2 test statistic. For example, consider the two χ^2 cell contributions for digit 2. They are:

$$Chi2Cell(d_2) = \frac{(165.43-160.7)^2}{160.7} = \frac{(156-160.7)^2}{160.7} = 0.14 \quad (7)$$

This is the case for the Pearson formulation of the expectation in the benchmarking case as the expectation is the mid-point between the sum of the two counting values. Consider now the other logical choice for the calculation of the χ^2 the benchmarking projection as the expectation.

3.1.3 The Direct Benchmark Projection as the Expectation

In the above section we assumed that the benchmark projection was a data sample and so we had two datasets with different digital frequencies each of which had the same number of total observations; this is the **Pearson** formulation. IF however, we elect to use the following **N-B profile** as the *expectation* for the i^{th} digit, d_i , projection:

$$E_i = \text{LOG}_{10} (1 + 1/d_i) \times Ni: 1, - - -, 9 \quad (8)$$

Where: N is the total number of realizations for the datasets under consideration—that is the number of values in the Observed dataset.

or the other possibility the **BPP** as the *expectation* for the i^{th} digit projection we have:

$$E_i = BPP_i \times Ni: 1, - - -, 9 \quad (9)$$

then, the j^{th} index is no longer needed and therefore the number in the summation will have nine elements and the χ^2 can be re-formed as:

[The N-B Log₁₀ Digital Frequency Profile]

$$\chi^2 = \sum_{i=1}^9 [(O_i - (\text{LOG}_{10} (1 + \frac{1}{d_i}) \times N))^2 / (\text{LOG}_{10} (1 + \frac{1}{d_i}) \times N)], df = 8. \quad (10)$$

[Benford Practical Profile]

$$\chi^2 = \sum_{i=1}^9 [(O_i - (BPP_i \times N))^2 / ((BPP_i \times N))], df = 8. \quad (11)$$

The implication for the magnitude of the χ^2 under the two different election assumptions: the **Pearson** and **Direct Benchmarking Projection** are very important. It is always the case that the χ^2 computed using the *Direct Benchmarking Projection* as the expectation is greater in magnitude than the χ^2 computed using the *Pearson Formulation*. Lusk and Halperin (2014a) show that the difference can be on the order of a 50% or so increase in the value of the χ^2 . This difference may materially affect the p-value that the auditor uses for inference for deciding on the use of extended procedures; we will illustrate this subsequently. Consider now the second issue relative to using the χ^2 test for extended procedures: The sample size issue.

3.2 Sample Size Anomaly

Cho and Gaines (2007, p.220), and just about all others, for example (Reddy and Sabastin (2012)) who have explored the χ^2 test as a differential DFP-expectation signal note:

“ χ^2 tests are very sensitive to sample size, having enormous power for large N, so that even quite small differences will be statistically significant. This test appears to be too rigid to assess goodness-of-fit well, especially since the Benford proportions do not represent a true distribution that one would expect to occur in the limit.”

The issue that Cho and Gaines identify leads to what is referred to, in the audit context, as the FPE-jeopardy. This means that because the width of the fail-to-reject-region decreases as the root of increases in the sample size, then large sample sizes create such a small fail-to-reject-region that very small differences relative to the Null provide p-value evidence that the difference is not due to random sampling chance, and so provide a justification for rejecting the Null of no difference. In the audit context the FPE-jeopardy can have onerous consequences as rejecting the Null suggests that the auditor should use extended procedures to investigate the account under audit as there is a p-value indication that the account materially differs from expectation. This is very expensive. If such unwarranted investigations are done on a regular basis the cost of the audit will not be calibrated or “in-sync” with the actual/realistic audit risk level. Simply said: *The actual audit risk level of the engagement should warrant/justify a certain number of extended procedure investigations; where the FPE-jeopardy is anonymously in play, the number of extended procedures investigations will be much larger than needed.* After some short period of time, the CPA LLP

will lose clients due to the cost of unwarranted investigations. This sample size “critique” of the χ^2 test discussed by Cho and Gaines is generally true in almost all statistical testing. In this regard, we suggest that the issue is not to reject the χ^2 screen in Digital Frequency Testing, but to calibrate the sample size so that this ideal test, the χ^2 , can be used in the audit context. In this regard, Lusk and Halperin (2014a) provide a validated sample size recalibration interval for sample accruals that are large and so would avoid the FPE-jeopardy. They recommend for samples in excess of 1000 observations that samples in the range [315 to 440] have desirable false positive error [*using extended procedures when an investigation is not warranted*] and false negative error [*failing to use extended procedures when an investigation is warranted*] profiles; we will use this sample size calibration in our investigations of the various test χ^2 possibilities.

4. The Menu of Choices for Using the χ^2 as a Screening Test for Signaling the Need to Use Extended Procedures

Recall that the critical aspect of this paper addresses the important question: *In the certification audit what is a reasonable way to decide if a particular account under audit should be “further” investigated using extended procedures?* This decision is critical in that using extended procedures are usually needed if the risk profile of a particular account seems relatively high, and the auditor must take the investigative steps to determine if the account could contain material error. This is of course the auditor’s charge from the SEC, PCAOB[AS 5] and the AICPA. The other side of the issue, as suggested above, is that extended procedures are very costly which makes the cost of the audit unrealistically high. The auditor seeks a balance between the FPEs and the FNEs relative to keeping the cost consistent with the risk level of the audit relative to the use of extended procedures.

Consider now the context for the various modeling possibilities where the χ^2 Digital Frequency Profile [DFP] test is a screening tool for extended procedures. Following, we will present:

1. recommendations that we have made to the public accounting firm to which we are the academic consultant and also have used in our Audit and Assurance course the text of which is Beasley, Elder & Arens (2012). Regarding these suggested DFP testing protocol alternatives, it is important to note, that there is no “correct” alternative; all are defensible and so possible and therefore a mix of the various possibilities is also plausible.
2. illustrative examples of six datasets, Appendix A, three of which have been reported in the literature as being representative of datasets that conform to the N-B profile and three of which are datasets that have been found to not conform to the N-B profile.
3. two aspects of the χ^2 distribution: (i) the overall χ^2 inferential test which will have uniformly eight (8) degrees of freedom and (ii) the χ^2 cell contribution. For the former, we recommend, and will use in the following sections of this paper, the 95% confidence level as the inferential-cut-point for considering the application of extended procedures. This seems consistent with the various Audit and Assurance texts that have presented the “acceptable” audit risk at 5% or [1-95%]. Other risk levels can of course be used. Ramos (2008) for example uses discovery sampling tables that range from a confidence level of 20% to 0.5%.
4. the Tamhane and Dunlop [TD] (2000, p.324) individual χ^2 cell value heuristic. They suggest that individual cell values are important signals of specific inherent variation from expectation. This can aid the auditor in focusing the investigation. The TD heuristic is: *Any χ^2 cell contribution greater than 1.0 is of interest as an indicator or signal of an important variance of **expectation from actual**.* We will be using this heuristic on a-cell-by-cell basis consistent with their recommendation as it logically focuses on the particular digits that are likely candidates for investigation over the two datasets. To be clear, there is NO statistical inference attached to this TD-signaling protocol—it is their heuristic. What still governs is the overall ; *this is the only statistically-based inference signal that can be used.* However, one can certainly use the TD criterion to form an opinion even in the case where the overall χ^2 does suggest an investigation. Recall, the FPE is the chance that the auditor incorrectly rejects the Null—i.e., that the multi-dimensional observed categorical profiles come from two identical profile generating processes and so investigates when it is not needed. Specifically, where the overall χ^2 produces a low p-value respecting the Null, then there is more of a justification for the auditor to be alert to a possible variance that needs to be investigated. In this regard, the menu of usual investigation choices will be noted in ***bold-italics***; they are:
 - i. When the overall χ^2 is associated with a p-value consistent with a FPE lower than 5%, then the account in question should create an audit action to: ***(i) query management and (ii) investigate the account under audit using extended procedures using the TD signals to focus attention on testing***

possibilities. This information should be recorded and later evaluated to determine if the investigation was in fact needed. This is a check, *à la Gosset*, of the adequacy of the FPE calibration (Note 3).

- ii. When the overall χ^2 is not associated with a low p-value, **then the TD signals should be recorded as part of the current audit record and used as a reference profile.** If it is later determined that the account or account groupings were in fact in need of extended procedure investigations, then this information can be used as a check on the False Negative Error [FNE] calibration. Finally,
5. selected aspects of the Decision Support System [DSS] called: Chi2DSS used to create most of this audit DFT information. The Chi2DSS is an Excel VBA open-access DSS. The Chi2DSS is available from the corresponding author without charge as an email download without restriction on its use.

For the various demonstrations we will be using the six datasets presented in Appendix A. We have decided to use these datasets as the agreement with the public accounting firm LLP to which we are the academic consultant prohibits the publication of the datasets or the results of the inferential analysis from data of their clients. Consider now the three modeling DFT possibilities in the audit context.

4.1 Model A: Two Samples of Different Accounts using the Pearson Formulation

This is a situation where the auditor wishes to test two accounts taken as random samples from:

Condition I Two different accounts in the same audit year,

Condition II One account taken from the current audit and one account from:

- a. a prior year where this account was tested and found to be free from material error, or
- b. a random sample from a reliable industry profile to be used as the benchmark.

4.1.1 Condition 1

For purposes of illustration (Note 4), we have the following datasets from Appendix A: GIC(25) A/P, n = 850 and GIC(25) Cash, n = 845

Table 4. Cell χ^2 values for the Comparison of the Samples: A/P and Cash Chi2DSS [**Tab: ComputationsTwoDataSetsOnly**]

First Digit Array	A/P Frequencies	GIC(25) Values	A/P χ^2 Cell	Cash Frequencies	GIC(25) χ^2 Cell
Digit 1	0.308	0.0289	0.302	0.0291	
Digit 2	0.184	0.0070	0.181	0.0071	
Digit 3	0.115	0.0009	0.116	0.0009	
Digit 4	0.107	1.0489	0.085	1.0551	
Digit 5	0.079	0.1050	0.085	0.1056	
Digit 6	0.058	0.0537	0.062	0.0540	
Digit 7	0.052	0.2781	0.060	0.2798	
Digit 8	0.052	0.0004	0.052	0.0004	
Digit 9	0.046	0.4910	0.057	0.4939	
Totals	850	2.014	845	2.026	

In this case, where the auditor compares two datasets, usually with different numbers of observations (Note 5) from the same audit year, the DFP question is: *Are these first digital frequencies between the two datasets different with respect to one another over the nine first digits?* The information generated by the Chi2DSS as presented in Table 4 indicates that overall the datasets are not different as the overall χ^2 is: 4.04 [2.014 + 2.026] which is far below the suggested 95% [FPE of 5%] cut-off of 15.507; however, there is a TD-indication that the number “4” is in fact different between the two tested datasets. Therefore, for the two sample information in Table 4, the audit-action-indication is: **As there is NO overall χ^2 difference between the two datasets at less than the 5% FPE, then there are no indications for an inquiry or extended procedures and the auditor would record this TD heuristic information for future reference.**

4.1.2 Condition II

Consider now the second condition where there is a download from a prior audit year or from the industry group as an audit benchmark; here we will use the same *GIC A/P dataset as the industry dataset* and for the Audit Account we will assume that it follows the frequencies of the *non-conforming dataset* due to Cho & Gaines (2007) [C&G] from 1995/96. We will assume, conservatively, that we have taken a random sample of 440 from the audit account as the number of observations in the audit datasets is 11 108 which is much greater than 1000.

Table 5. Comparison of an Account Under Audit [using C&G 1995/96 dataset] with the Industry Profile [using the GIC A/P dataset] Chi2DSS[*Tab: FreqEQ2andE10*]

First Digit Array	Assumed Industry Profile[GIC]	Industry χ^2 Values	Cell	Assumed Audit Data [C&G:95/96]	Audit Cell χ^2 Values
Digit 1	0.308	1.40		0.244	2.71
Digit 2	0.184	0.56		0.217	1.08
Digit 3	0.115	1.42		0.158	2.74
Digit 4	0.107	0.11		0.096	0.22
Digit 5	0.079	0.61		0.102	1.17
Digit 6	0.058	0.04		0.063	0.08
Digit 7	0.052	0.03		0.048	0.06
Digit 8	0.052	0.87		0.032	1.68
Digit 9	0.046	0.08		0.040	0.15
Totals	850	5.12		440	9.91

In this case, there is not an overall χ^2 indication that the comparative result is different at the 95% confidence level as the χ^2 value is 15.03 [5.12 + 9.91] which is less than 15.507. However, there are a number of χ^2 cell TD-differences. There is a difference for both datasets for the “1s” and the “3s” and for the audit dataset the “2s”, “5s”, and “8s” are marginally different from the industry benchmark dataset. As the overall value is 15.03 which is close to but less than 15.507, then the audit-action would be: ***This profile of digital differences would be recorded for future reference.*** If the overall χ^2 had been greater than 15.507 and we had the same five χ^2 cell differences, then this likely would have required the auditor to ***make an inquiry and use extended audit procedures to collect evidence focused on the TD signals for the information offered by management.***

As a final note for the information profiled in Table 5, there is an interesting aspect of this example that we wish to note. Recall that the auditor took a random sample of 440 as the actual sample had 11 108 observations. Let us assume that the auditor in fact used this large actual sample size and we had the same profile as is presented in Table 5. In this case, six χ^2 cell contributions are greater than 1.0 and the overall χ^2 is 35.2! This is a very different result from that of Table 5 and is due solely to the large sample size, which is why one must take a sample when the number of observations is large.

4.2 Model B: Two Samples: One an Actual Account and the other a Benchmark Projection using the Pearson Formulation.

Here the sample sizes will be the same, as the benchmark is a projection as opposed to a separate accrued sample. There can be two such projections: the BPP or the Log₁₀. In this case then, the auditor has a choice between the two expectations that will be used to generate the benchmark. We will examine both possibilities; in an actual audit context we recommend using only one of the benchmarks and for consistency purposes to NOT change this benchmarking context. Let us assume that we have *as the audit dataset* the non-conforming dataset from R&S 91/92. As the R&S 91/92 dataset has 11 661, observations assume we took a random sample in the interval [315 and 440] and will use 378 as the sample size. Further, let us assume that the BPP Projection is the other sampled dataset to be used as the benchmark.

Table 6. Audit Benchmark Comparison Using the BPP Projection and the Pearson Formulation Chi2DSS[*Tab:EQ2FrequenciesOnly*]

First Digit Array	Assumed Dataset	Audit	R&S Cell χ^2 Values	BPP Projections	BPP Cell χ^2 Values
	R&S 91/92				
Digit 1	0.234		1.12	0.289 19	1.12
Digit 2	0.180		0.11	0.194 62	0.11
Digit 3	0.152		0.43	0.126 65	0.43
Digit 4	0.107		0.25	0.090 61	0.25
Digit 5	0.102		0.74	0.075 44	0.74
Digit 6	0.097		1.24	0.064 31	1.24
Digit 7	0.072		0.48	0.054 08	0.48
Digit 8	0.034		0.93	0.054 87	0.93
Digit 9	0.024		1.79	0.050 52	1.79
Totals	378		7.09	378	7.10

Here we see that the overall χ^2 is 14.19 [7.09 + 7.10] and is not at the 95% cutoff. Digits “1”, “6” and “9” are identified by the TD criterion. ***In this case, an inquiry is not warranted consistent with the audit action protocol that we have adopted; however, the TD-signal would be recorded for future reference.***

Now consider the same audit dataset however, this time benchmarked using the Log₁₀ projection. These comparative results are presented in Table 7.

Table 7. Audit Benchmark Comparison Using the Log₁₀ Projection and the Pearson Formulation Chi2DSS[*Tab:EQ2FrequenciesOnly*]

First Digit Array	Assumed Dataset	Audit	R&S Cell χ^2 values	Log10 Projections	Log10 Values χ^2
	R&S 91/92				
Digit 1	0.234		1.61	0.301 03	1.61
Digit 2	0.180		0.01	0.176 09	0.01
Digit 3	0.152		0.49	0.124 94	0.49
Digit 4	0.107		0.09	0.096 91	0.09
Digit 5	0.102		0.53	0.079 18	0.54
Digit 6	0.097		1.03	0.066 95	1.03
Digit 7	0.072		0.28	0.057 99	0.28
Digit 8	0.034		0.66	0.051 15	0.66
Digit 9	0.024		1.29	0.045 76	1.29
Totals	378		5.99	378	6.00

Here we see the effect of changing the screening projection. In this case, the information is basically the same as the overall χ^2 is 11.99 [5.99 + 6.00] still under the 95% cut-off and the same three digits “1”, “6” and “9” are identified as the comparative TD-signal. This accounting, anecdotal though it is, shows what we find is the typical case that: Varying the benchmark usually does not change the nature of the reaction of the auditor; ***in both cases [Tables 6 & 7] the TD-information is simply recorded and reviewed as more DFT information is collected.***

4.3 Model C The Benchmark as the Expectation

The last point of demonstration is to use *the benchmark as the expectation* as presented in EQ10 or EQ11. Here also there are two cases that are relevant for benchmarking: the BPP and the Log₁₀. In the first case, we will use the Hill (1998) non-conforming data as the Account under audit and the BPP as the benchmark projection; and for the second case the Nigrini (1996) conforming data and the Log₁₀ as the expectation will be used.

The two Tables: 8 & 9 produced in these cases are:

Table 8. The Hill (1998) Data Non-Conforming Data as the Assumed Audit Dataset Compared using the BPP Projection as the Expectation Chi2DSS[*Tab: FreqEQ2andE10*]

First Digit Array	Assumed Audit Dataset [Hill]	BPP Projections	Cell Values	χ^2
Digit 1	0.147	0.289 19	51.94	
Digit 2	0.100	0.194 62	34.18	
Digit 3	0.104	0.126 65	3.01	
Digit 4	0.133	0.090 61	14.73	
Digit 5	0.097	0.075 44	4.58	
Digit 6	0.157	0.064 31	99.26	
Digit 7	0.120	0.054 08	59.70	
Digit 8	0.084	0.054 87	11.49	
Digit 9	0.058	0.050 52	0.82	
Totals	743	743	279.72	

Here we see that if the Audit datasets followed the Hill profile and we used the BPP as the projection expectation that there is a clear signal of a difference as that the overall χ^2 is very large compared to the 95% confidence level of 15.507. Here the p-value for the FPE is < 0.0001 strongly suggesting that there are DFP differences. As this is a non-conforming dataset, this is the correct decision. Also, eight of the nine χ^2 cell TD-contributions are > 1.0 . This DFT profile *requires queries and follow-up testing using extended procedures*. Failing to do so the auditor likely risks to be judged as not being attentive to best-practices' inferential common sense. As a further comparison of interest, IF we had used the BPP as another dataset in the Pearson modality, then the audit alert profile would be very similar. Specifically, in the Pearson case, the overall χ^2 is greater than 100 and so also much greater than the critical investigation cut-off of 15.507 and seven of the nine χ^2 cell TD-contributions are > 1.0 . In this case then *queries and extended procedures are also called for*.

For the next comparison, we will use the Nigrini (1996) conforming dataset and the Log₁₀ projection. As this sample has 78 640 observations, the auditor should take a random sample. Assume that we take 315.

Table 9 The Nigrini88 (1996) Conforming Data Compared using the Log₁₀ Projection as the Expectation Chi2DSS[*Tab: FreqEQ2andE10*]

First Digit Array	Assumed Audit Data [Nigrini88]	Log10 Projections	Cell Values	χ^2
Digit 1	0.306	0.301 03	0.03	
Digit 2	0.178	0.176 09	0.01	
Digit 3	0.127	0.124 94	0.01	
Digit 4	0.095	0.096 91	0.01	
Digit 5	0.078	0.079 18	0.01	
Digit 6	0.065	0.066 95	0.02	
Digit 7	0.056	0.057 99	0.02	
Digit 8	0.050	0.051 15	0.01	
Digit 9	0.045	0.045 76	0.00	
Totals	315	315	0.11	

In this case *there are no indications either in terms of the overall χ^2 or from the TD signals that extended procedures or inquiries would be warranted*. As the Nigrini88 data was a conforming dataset this is the correct decision. However, as an extension, let us assume that we used the sample size of the Nigrini88 data, 78 640 while

holding all other indications constant. In this benchmarking case, the overall χ^2 is 28.0 larger than the 95% confidence level of 15.507 and eight of the χ^2 cell TD-contributions are > 1.0. In this case, ***the protocol suggests an inquiry and extended procedures testing***. This is not the correct decision as the Nigrini88 data is a conforming dataset and the low FPE is solely due to the large sample size.

5. Summary, A Caution, and Future Developments

5.1 Summary

We have offered three modeling perspectives that have relevance in the audit context regarding the use of Newcomb-Benford digital frequency testing. To summarize the various possibilities we have prepared a taxonomy which can be used as a reference in executing the DFT protocols.

Table 10. The Taxonomy of Information regarding DFT using the χ^2 as the Inference Model

Sampling Protocol	FPE Null for Inference	Benchmark	Computing Method	Sample Size	Overall χ^2 Inference	Cell χ^2 TD-Inference
Two samples from one population generating process.	The two samples are categorically identical to each other.	The two samples are benchmarked by their marginal values.	Pearson EQ2 only	A random sample in the range [315 to 440] is suggested if the account has more than 1000 items.	FPE < 5% warrants audit inquiry <u>and</u> extended procedures. FPE \geq 5% no audit action only recording.	Cell χ^2 > 1.0 should be recorded and used as comparative information or to focus audit inquiries.
One Sample from the Current Audit	The categorical profile of the audit sample is identical to that of the Benchmark	BPP or N-B Log_{10}	Pearson EQ2 or [EQ10 or EQ11]	Same as above	Same as above	Same as above

5.2 Caution

The caution is simply that DFT using the χ^2 as the inference model is dependent on a reasonable sample size. Recall the very different results as demonstrated by the Nigrini88 conforming data in Table 9 where the sample size of the audit account was 78 640. This is a simple but clear indication that large sample sizes are not only a risk factor—they **guarantee FPE anomalies—i.e., committing unwarranted investigations and compromising the effective use of audit resources**. Another caution is that using the benchmark as the expectation may result in the overall χ^2 effectively doubling compared to the value of the χ^2 value using the Pearson two-sample computational form.

To illustrate this let us re-consider the information in Table 5 as Table 11 following:

Table 11. Comparison of an Account Under Audit [using C&G 1995/96 dataset] with the Industry Profile [using the GIC A/P dataset]

First Digit Array	Assumed Industry Profile[GIC]	Assumed Audit Data [C&G:95/96]
Digit 1	0.308	0.244
Digit 2	0.184	0.217
Digit 3	0.115	0.158
Digit 4	0.107	0.096
Digit 5	0.079	0.102
Digit 6	0.058	0.063
Digit 7	0.052	0.048
Digit 8	0.052	0.032
Digit 9	0.046	0.040
Totals	440	440

We have modified the sample size to 440 in both cases to allow the comparisons between the Pearson and the Benchmark. Using the Chi2DSS [*Tab: FreqEQ2andE10*] for the computations the Pearson computation of EQ2 is $\chi^2 = 11.4$ which is less than the 95% FPE cut-point of 15.507 and there were three TD signals [Digits; 1, 3 & 8]; the audit action that this suggests *is only recording the TD signals*. However, if the auditor believed that the audit dataset should be benchmarked against the industry dataset, then the overall χ^2 is 23.8 which is greater than 15.507 the 95% cut-off and there were five TD signals [Digits; 1, 2, 3, 5 & 8]. In this case, *inquiry and extended procedures are likely in order*. These are VERY different results; *they are due solely to the selection of the expectation*.

Which expectation formulation is used depends on the belief underlying the state of nature for DFT. The DFT-researchers cited in the references have used the Log_{10} as the point process expectation and so used EQ10 in their calculation protocol. This simply means that they are suggesting a test against a theoretical set of point values. In this format, the theoretical set of point values is fixed with NO variation—it is the population parameter set not a sampling estimate of it. Whereas, in the Pearson formulation, both are samples assumed to be drawn from a generation process [usually assumed to be the same for both sample sets] and therefore both have variation around the theoretical set of point values of the unknown generating process of the population profile point set. This is the reason that the χ^2 value for EQ10 [or EQ11] is always greater than that using EQ2 as the precisions are very different in the calculation assumptions, and as a result the p-values are also different. This is the same idea as testing a sample mean, \bar{x} , against μ as opposed to testing \bar{x}_A from process A and \bar{x}_B from Process B where both \bar{x}_A and \bar{x}_B are samples from the respective processes. In our view, we like to think that the benchmark projection is just another generating process really invented by Newcomb and popularized by Benford. In this context, we advise the auditor to use either the Corrected Benford set of sample means or the Log_{10} profile as a sample from a generating process and not the absolute gold-standard categorical population point process that has no variation.

5.3 Future Developments

What is needed is a DFT archive where Conforming and Non-conforming **audit datasets** are stored and available to researchers in this interesting and pertinent audit best practices area. In this exposition, we have used six datasets that are available in the literature to demonstrate the various testing protocols that are possible in the audit context. Having actual audit datasets that *conform* and do *not conform* will be the only way to ferret out the practical FPE and FNE issues for DFT in the audit context. Further, such an archive will enable the development of calibrated FPEs and the FNEs in the inferential context of the audit with respect to the TD-signals. To aid in this endeavor, we invite those interested in developing such a DFT-archive to send to the corresponding author datasets and they will be posted on our “commons” link which is publically available free of charge.

Acknowledgments

We wish to thank Dr. H. Wright, *Boston University*: Department of Mathematics and Statistics, the participants at the Research Workshop at the *State University of New York: Plattsburgh* for their detailed and most helpful suggestions, in particular Dr. Héroux. Finally, we wish to thank Mr. Nicolae Lungu and Mr. Tom Stretton, CPA: *McSoley and McCoy LLP*, [<http://www.cpavt.com/>] for their council and support in the development of this project.

References

- Beasley, M., Elder, R., & Arens, A. (2012). *Auditing and assurance services* (14th ed.). Pearson Publishing, New York, NY USA: ISBN: 13-978-013-257609-3
- Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78, 551-572.
- Cho, W.K.T., & Gaines, B.J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *American Statistician*, 61, 218-223. <http://dx.doi.org/10.1198/000313007X223496>
- Chatterjee, S.K. (2003). *Statistical thought: A perspective and history* (1st ed.). Oxford University Press, Oxford; NY USA ISBN: 01-985-5253-11. <http://dx.doi.org/10.1093/acprof:oso/9780198525318.001.0001>
- Geyer, C., & Williamson-Pepple, P. (2004). Detecting fraud in data sets using Benford's law. *Communications in Statistics: Simulation and Computation*, 33, 229-246. <http://dx.doi.org/10.1081/SAC-120028442>
- Hickman, M., & Rice, S. (2010). Digital analysis of crime statistics: Does crime conform to Benford's law? *Journal of Quantitative Criminology*, 26, 333-349. <http://dx.doi.org/10.1007/s10940-010-9094-6>
- Hill, T. P. (1998). The first digit phenomenon. *American Scientist*, 86, 358–363. <http://dx.doi.org/10.1511/1998.4.358>

- Lee, C-H., Lusk, E.J. & Halperin, M. (2014). Content analysis for detection of reporting irregularities: Evidence from restatements during the SOX-Era. *Journal of Forensic and Investigative Accounting*, 6, 99-122.
- Ley, E. (1996). On the peculiar distribution of the U.S. stock indexes' digits. *American Statistician*, 50, 311-313. <http://dx.doi.org/10.1080/00031305.1996.10473558>
- Lusk, E.J. & Halperin, M. (2014a). Calibrating the χ^2 Test for use in Digit Frequency Testing: A Note. *International Business Research*, 7, 134-147.
- Lusk, E.J. & Halperin, M. (2014b). Using the Benford Datasets and the Reddy & Sebastin Results to Form an Audit Alert Screening Heuristic: A Note, *IUP Journal of Accounting Research and Audit Practices*, in Press.
- Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4, 30-40. <http://dx.doi.org/10.2307/2369148>
- Nigrini, M. (1996). A taxpayer compliance application of Benford's law. *Journal of American Taxation Association*, 18, 72-91.
- Nigrini, M J (1999). I've got your number. *Journal of Accountancy*, 187, 79-83.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series*, 50, 157-175. <http://dx.doi.org/10.1080/14786440009463897>
- Public Company Accounting Oversight Board [PCAOB] 2007: Auditing standard No. 5 [AS 5] Release No. 2007-005 24 May 2007 <http://pcaobus.org/Standards/Pages/default.aspx> Accessed 16 Oct 2013. (Accessed 12 March 2014).
- Ramos, M. (2008). *Practitioners guide to GAAS:2008* (7th ed.). J. Wiley and Sons, New York, NY USA: ISBN: 978-0-470-13531-0.
- Reddy Y.V., & Sebastin, A. (2012). Entropic analysis in financial forensics. *The IUP Journal of Accounting Research and Audit Practices*, 11, 42-57.
- Report 8 Oct 2002: *Delivered to the senate committee on governmental affairs: Financial oversight of Enron: The SEC and private-sector watchdogs* <http://www.gpo.gov/fdsys/pkg/CPRT-107SPRT82147/pdf/CPRT-107SPRT82147.pdf> (Accessed 8 April 2014).
- Tamhane, A. & Dunlop, D. (2000). *Statistics and data analysis*. (1st ed.). Prentice Hall, Upper Saddle River, NJ USA: ISBN: 0-13-744426-5.

Notes

Note 1. See for example: Nigrini (1996 & 1999); Ley (1996); Hill (1998); Geyer & Williamson-Pepple (2004); Cho & Gaines (2007); Hickman & Rice (2010); & Reddy & Sebastin (2012).

Note 2. These expectations are the averages of the Pearson Expectations as the total for the Projection is slightly larger.

Note 3. Interestingly, over-sampling is simply illustrated by the screening effect of the t-distribution as initially published by Gosset under the name of Student [and subsequently refined by R.A. Fisher]. The confidence interval of the t-distribution is slightly wider than the distribution that does not account for the sampling effect in forming the sampling distribution. In this case, the correct t-distribution, calibrated using the degrees of freedom, used to form the FPE-screen relatively will signal fewer investigations as the acceptance [fail to reject the Null] region is larger. See for more information: Chatterjee (2003).

Note 4. In all of the illustrations to follow we will use datasets that are available from the literature or as downloads. This will facilitate the pedagogy if instructors wish to use these models in the classroom.

Note 5. When the sample sizes are different, then the ratio relationship between the overall using the Pearson EQ2 formulation and that using the Benchmark [Eq10 or 11] as the expectation is no longer bounded in the range [(0 to 1), excluding the 50% point] as it is when there is one dataset and the projection uses the exact number of observations of the audit dataset.

Appendix A Datasets Used as Illustrations*

Indices	Digit 1	Digit 2	Digit 3	Digit 4	Digit 5	Digit 6	Digit 7	Digit 8	Digit 9
<i>N-BLog10</i>	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046
C&G95/96	0.244	0.217	0.158	0.096	0.102	0.063	0.048	0.032	0.040
R&S91/92	0.234	0.180	0.152	0.107	0.102	0.097	0.072	0.034	0.024
Hill	0.147	0.100	0.104	0.133	0.097	0.157	0.120	0.084	0.058
GIC:A/Pay	0.308	0.184	0.115	0.107	0.079	0.058	0.052	0.052	0.046
GIC:Cash	0.302	0.181	0.116	0.085	0.085	0.062	0.060	0.052	0.057
Nigrini88Tax	0.306	0.178	0.127	0.095	0.078	0.065	0.056	0.050	0.045

*Sample sizes in respective order: N/A; 11 108; 11 661; 743; 850; 845; 78 640.

The first row, in ***Bold-Italics***, presents the N-B theoretical digital frequencies that are the classic benchmark frequencies. The next three entries as **bolded** are *non-conforming* datasets. *These are NOT audit datasets that have been tested and found to be materially in error by using extended procedures.* They are, however, datasets that others have argued as not conforming to the N-B Log₁₀ profile. In Rows 2 & 3 are datasets taken from the Federal Election Commission [FEC] reported Committee to Committee In-Kind contributions. Row 2 is provided by Cho and Gaines [C&G] for the year 1995/96 (2007, Table 1: p.218). Row 3 is due to Reddy and Sebastin [R&S] (2012, Appendix 4:p. 56/57). Row 4 is the data published by Hill (1998, Figure 5, p.363) where he asked 743 first-year undergraduate students to write down a six-digit number. The last three datasets are **shaded**; they are expected to *conform* to the N-B Log₁₀ Profile. *These are also NOT audit datasets that have been tested and found to be free from material error.* In Rows 5 and 6 are two datasets downloaded from the GIC Sector 25 using the WRDS™ database of the Wharton School of firms listed on market trading exchanges and so are required to have a certified audit suggesting by extension that the data was free from material error. The first of these is reported *Accounts Payable* and the second is *Cash Balances* both at Year-end. We selected *Accounts Payable* as Nigrini (1999) also selected this type of account for analysis. Row 7 is taken from Nigrini (1996, Table 2:p.80) which are the first digits of Interest Received as reported on filed individual US Federal Income Tax returns for the year 1988.